

Open Research Online

The Open University's repository of research publications and other research outputs

Summarising News Stories for Children

Conference or Workshop Item

How to cite:

Macdonald, Iain and Siddharthan, Advaith (2016). Summarising News Stories for Children. In: Proceedings of the 9th International Conference on Natural Language Generation (INLG 2016), ACL, pp. 1–10.

For guidance on citations see [FAQs](#).

© 2016 Association for Computational Linguistics



<https://creativecommons.org/licenses/by-nc-nd/4.0/>

Version: Version of Record

Link(s) to article on publisher's website:

<http://dx.doi.org/doi:10.18653/v1/w16-6601>

<https://www.aclweb.org/anthology/W/W16/W16-6601.pdf>

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online's data [policy](#) on reuse of materials please consult the policies page.

oro.open.ac.uk

Summarising News Stories for Children

Iain Macdonald

Computing Science
University of Aberdeen
Scotland, U.K.

`iain.j.macdonald.12@aberddeen.ac.uk`

Advaith Siddharthan

Computing Science
University of Aberdeen
Scotland, U.K.

`advaith@abdn.ac.uk`

Abstract

This paper proposes a system to automatically summarise news articles in a manner suitable for children by deriving and combining statistical ratings for how important, positively oriented and easy to read each sentence is. Our results demonstrate that this approach succeeds in generating summaries that are suitable for children, and that there is further scope for combining this extractive approach with abstractive methods used in text simplification.

1 Introduction

Automatic text summarisation is a research area with half a century of history, with Luhn (1958) discussing as far back as 1958 the task he called “auto-abstracting of documents”. This field has evolved considerably with a large number of unsupervised and supervised techniques for summarising documents reported in the literature (see Nenkova and McKeown (2012) for an overview). The vast majority of publications focus on sentence selection based on notions of information content and topicality; such methods are referred to collectively as extractive summarisation. We adapt one such well understood notion of informativeness to incorporate other desirable characteristics such as how positive or optimistic sentences are and how difficult they are to read, with the goal of generating news summaries that are suitable for children.

We are targeting a similar demographic of children as that of the British Broadcasting Cor-

poration’s CBBC Newsround¹, a television programme and website dedicated to providing children in the age range of 6–12 years with news suitable for them (Newsround, 2011). This is primarily motivated by two factors: the importance of young people engaging with current affairs and the potential benefits of automating the creation of children’s news articles.

Multiple studies have highlighted potential links between youth civic engagement (defined by Adler and Goggin (2005) as active participation in the life of a community to improve conditions for others or to help shape the community’s future) with the use of various forms of news media (see Boyd and Dobrow (2010) for a good overview). However, while children’s news sources exist, possibly the best known being the aforementioned Newsround, they are time consuming to maintain, and as a result very few news stories are made available through them. For instance, Newsround has only six journalists working to maintain the website (Newsround, 2008), who focus more on multimedia content, so only around five articles a day are published for children. While the guidelines used to produce Newsround articles are not public, we have observed that they are shorter than those that appear on the main news site, use simpler language, and also try to stay upbeat, avoiding upsetting news where possible. Our primary objective is to automate the generation of such news stories for children using an extractive approach, though the further potential for abstractive ap-

¹<http://www.bbc.co.uk/newsround>

proaches such as text simplification is also discussed. In order to achieve this objective, there are four key components described in this paper:

- A measure of how informative a sentence is.
- A measure of how positive or negative a sentence is.
- A measure of how difficult a sentence is to read and understand.
- A formula for combining the combining the previous measures.

We describe these components and our evaluation methodology in §2 and our results in §3 before discussing our contributions with respect to related work in §4 and presenting our conclusions in §5.

2 Method

We based our summariser on SumBasic, a contemporary summariser that has been shown to perform well in evaluations in the news domain (Nenkova and Vanderwende, 2005) and is easy to adapt. SumBasic is a greedy algorithm that incrementally selects sentences to create a summary with a similar distribution of words as the input document(s). It begins by estimating the probability of seeing each word w_i in the input as $p_{input}(w_i) = n/N$, where n is the frequency of w_i in the input and N is the total number of words in the input. It then assigns a score to each sentence S_j which is the average probability of all the words in the sentence $Score_{SumBasic}(S_j) = \sum_{w_i \in S_j} p(w_i)/length(S_j)$. Sentences are selected in decreasing value of the score, and each time a sentence is incorporated in the summaries, the probabilities of words contained in the sentence are discounted to reduce the chance of selecting redundant sentences. We extended this algorithm to incorporate sentiment and ease of language as described below.

2.1 Information Score

We based our information metric on the SumBasic metric proposed by Nenkova and Vanderwende (2005):

$$Score_{SumBasic}(S_j) = \frac{1}{|\{w_i | w_i \in S_j\}|} \sum_{w_i \in S_j} p_{input}(w_i)$$

where the denominator denotes the number of words in the sentence. We adapted this metric in two ways:

1. A list *Stop* of 173 common stop words (University of Washington, 2012) was incorporated, and these were discounted in the calculations.
2. A peculiarity of news reporting in English is that the central information is often summarised within the first two sentences; this is sometimes referred to as the inverted pyramid structure, widely believed to have been developed in the 19th century (Pöttker, 2003), and the most common structure for print, broadcast and online news articles in English (Rich, 2015, p. 162). To account for this, we increased the score of the first sentence by a factor of 2 and the second by a factor of 1.5.

Our implemented information score is:

$$Score_{info}(S_j) = \frac{IPW}{|\{w_i | \begin{smallmatrix} w_i \in S_j \\ w_i \notin Stop \end{smallmatrix}\}|} \sum_{\begin{smallmatrix} w_i \in S_j \\ w_i \notin Stop \end{smallmatrix}} p_{input}(w_i)$$

where *IPW*, the inverted pyramid weight, is 2 for first sentence, 1.5 for second sentence and 1 otherwise.

2.2 Sentence Difficulty Score

Sentence difficulty is often assessed as some combination of lexical and syntactic difficulty. Typical heuristics such as readability formulae (Dale and Chall, 1948; Kincaid et al., 1975; Gunning, 1952; Mc Laughlin, 1969) are intended for scoring entire texts, rather than individual sentences. Alternately, psycholinguistic data for vocabulary such as the Bristol Norms (Stadthagen-Gonzalez and Davis, 2006; Gilhooly and Logie, 1980) exist for age of acquisition, familiarity, etc., but are relatively small (the Bristol Norms contain only 3,394 words).

To more directly assess linguistic suitability for children, we used a language model derived from historical BBC Newsround stories. Text-STAT (Hüning, 2002) was used to acquire 1000 Newsround URLs and ICEweb (Weisser, 2013) was used to extract the text from these web page. The probability of every word in the corpus was calculated, resulting in a lexicon of over

12,500 words. Lexical difficulty was then estimated in the same manner as importance in the section above; i.e. as the average probability of the words in the sentence, but this time according to the Newsround model. We excluded names from the calculation by matching words against a large collection of names (Ward, 1993):

$$Score_{lex}(S_j) = \frac{1}{|\{w_i | \begin{smallmatrix} w_i \in S_j \\ w_i \notin Names \end{smallmatrix}\}|} \sum_{\substack{w_i \in S_j \\ w_i \notin Names}} p_{newsround}(w_i)$$

We used a simple sentence length heuristic for syntactic difficulty, to give a combined difficulty score:

$$Score_{diff}(S_j) = \frac{Score_{lex}(S_j)}{|\{w_i | w_i \in S_j\}|}$$

2.3 Sentiment score

We implemented hybrid of a statistical and a rule based sentiment analysis component.

Supervised sentiment classifier: The statistical component was implemented as a supervised Naïve Bayes classifier with unigram, bigram and trigram features. We first experimented with training it on a large corpus of positive and negative movie reviews (Pang and Lee, 2004). We were however not satisfied with the quality of classifications for news stories. The key issue was the difference in vocabulary usage in the two genres; e.g. a word such as “terrifying” features prominently in positive movie reviews, but should not predict positive sentiment in a news story. For genre adaptation, a new dataset was created specifically for our purpose by taking a pre-existent dataset of 2,225 BBC articles assembled for topic classification (Greene and Cunningham, 2006). These articles were then manually labelled as positive, negative or neutral based just on the topic of the story, and the sentences from the positive and negative articles were added to the training data from the movie review dataset. This augmentation was observed to produce better results on new stories, but no formal evaluation was carried out on this particular aspect. For a sentence with n words, Naïve Bayes returns conditional prob-

abilities for each class (Pos and Neg), calculated as:

$$p(Pos|w_{1..n}) = p(Pos) \prod_{i=1..n} p(w_i|Pos)$$

$$p(Neg|w_{1..n}) = p(Neg) \prod_{i=1..n} p(w_i|Neg)$$

From these, we calculate a sentiment score as:

$$Score_{nb}(S_j) = \frac{p(Pos|w_{1..n})}{p(Pos|w_{1..n}) + p(Neg|w_{1..n})}$$

Dictionary based approach: In an effort to further overcome vocabulary issues with the statistical system, we also incorporated a dictionary-based approach. We used a sentiment dictionary with around 2000 positive and 4800 negative words respectively (Liu et al., 2005). The classifier simply starting with a sentiment score of 0.5 and incremented or decremented by 0.1 for every word in a sentence found in the positive or negative dictionary respectively.

$$Score_{dict}(S_j) = 0.5 + \sum_{\substack{w_i \in S_j \\ w_i \in Dict_{pos}}} 0.1 - \sum_{\substack{w_i \in S_j \\ w_i \in Dict_{neg}}} 0.1$$

2.4 Combining Scores

In order to combine scores, we first converted each individual score into its standard score (also called z-score); a renormalisation that gives each score a mean of 0 and a standard deviation of 1 over all sentences in the input. Following this step, a score for each sentence was computed as follows. First, the statistical and dictionary based (standardised) sentiment scores were combined in the ratio three is to one to give a single sentiment score:

$$Score_{sent}(S_j) = \frac{2}{3}Score_{nb}(S_j) + \frac{1}{3}Score_{dict}(S_j)$$

The final sentence score was then computed as a linear function of the scores for informativeness, difficulty and sentiment:

$$Score_{children}(S_j) = 5 \times Score_{info}(S_j) \\ + 5 \times Score_{sent}(S_j) \\ + 2 \times Score_{diff}(S_j)$$

The weightings were set by hand based on manual experimentation. We found that within a single news report, there was limited variation in sentence difficulty; this score could be assigned a higher weight in a multi-document summarisation task.

2.5 Experimental Setup

Evaluation platform: Amazon’s Mechanical Turk service² was utilised to create a survey to compare different summaries of news articles. Various studies have been carried out into the quality of data provided by Mechanical Turk with the general consensus of these seeming to be that, provided the questions are clear and that the instructions are intuitive, the data generated from Mechanical Turk is of a high quality (Ramsey et al., 2016; Buhrmester et al., 2011; Rand, 2012).

Summariser settings: We compared two summariser settings, the original SumBasic score for informativeness ($S_{SumBasic}$), and the other that combined informativeness with ease of reading and sentiment ($S_{children}$). For both settings, we set the required summary length to either one hundred words or half the length of the original article, whichever was smaller. With respect to how this was implemented in the iterative summariser described at the top of §2, any sentence that would cause the summary to exceed this length was ignored and the next highest rated sentence was given a chance in its place. Further, to prevent poorly scoring sentences being included, a minimum z-score limit was set to -0.25 below which sentences would be rejected. For both summarisers, sentences in the summary were reordered to correspond to their original ordering in the news article.

Evaluation data: We sampled 9 news articles to summarise, six from the BBC and one each from The Guardian, The Independent and Sky News. For the BBC articles, we generated a corpus of 1000 Newsround stories using TextSTAT (Hüning, 2002), and iteratively picked one using a random number generator, and then

checked that it was based on an article on the main BBC webpage (we did this in order to conduct a further comparison to the manually written Newsround story [c.f. §3.1]). The first six articles found to meet this criteria were used. An additional article was taken from each of The Guardian, The Independent and Sky News, again by sampling at random from a corpus of 1000 articles generated using TextSTAT.

These articles were then split into three surveys each with two BBC articles and one of the other three articles. For each article participants were presented with the two summaries produced by NSFC and GS, side by side, labelled ‘A’ and ‘B’, in a randomised order and without any information on how they were produced. They were provided a link to the original news report, but not forced to read it. Examples of summaries used in the evaluation are provided in Table 1. Participants were then asked to answer a four comparison questions on a five point scale [“A is significantly more X”, “A is slightly more X”, “Not sure, or equally X”, “B is slightly more X” and “B is significantly more X”], where X is the word in bold font in the questions below:

Q1 Which of these summaries is more **informative**?

Q2 Which of these summaries is more **positive**?

Q3 Which of these summaries is more **easy** to read and understand?

Q4 Overall, which of these summaries do you believe is more **suitable** for a child?

Finally, we asked a single non-comparison question for each summary on a five point scale [“Strongly disagree”, “Disagree”, “Not sure”, “Agree”, “Strongly Agree”]:

Q5 I would consider showing summary {A|B} to a child if I wanted them to know more about this news story.

Design: We solicited nine participants for each survey, twenty-seven in total, resulting in each question being answered eight-one times (twenty-seven participants, three articles each).

²<http://www.mturk.com/>

| NSFC | GS |
|--|---|
| A blaze that swept through a dogs' home has now claimed the lives of 60 animals, police have said. More than 150 dogs were rescued from the fire, which broke out at Manchester Dogs' Home in Moss Brook Road in Harpurhey on Thursday evening. Greater Manchester Fire and Rescue Service (GMFRS) tweeted its thanks to people who have donated money, saying: "One hundred and fifty dogs rescued. Thousands of pounds donated. Thank you Greater Manchester." The RSPCA described the fire as "heartbreaking". The Manchester home was established in 1893 and cares for more than 7,000 dogs every year. | The newspaper has also captured aerial footage showing the extent of the damage caused by the blaze. In the aftermath of the fire, the manager of the home said 60 dogs had been housed in the worst-affected building. Hundreds of messages of sympathy have been left on the JustGiving page, as the amount of money donated continues to rise. A number of people, including police officers and staff were quickly on the scene and put their life on the line to help with the rescue effort. The RSPCA described the fire as "heartbreaking". |
| Doctors have warned that almost half of all adults in Britain will be classified as obese within the next 20 years. They predict that on current trends an extra 11 million people will be severely overweight by 2030, bringing the total to 26 million. Only tough government action, including a tax on unhealthy food, can slow the trend, they say. At the top is a 10% tax on high-calorie food and drink. "People know obesity is a real problem. People don't know, as individuals, what to do about it." | The doctors have produced a league table of possible actions that could be taken to curb the epidemic. At the top is a 10% tax on high-calorie food and drink. "People know obesity is a real problem. People don't know, as individuals, what to do about it." "Governments do know what to do about it and if they could persuade people, as they easily could, it would be a popular action." Tam Fry, of the National Obesity Forum, said: "Children are born thin. It's what we do to children that makes them obese." |
| A new species of titanosaur unearthed in Argentina is the largest animal ever to walk the Earth, palaeontologists say. Based on its huge thigh bones, it was 40m (130ft) long and 20m (65ft) tall. A film crew from the BBC Natural History Unit was there to capture the moment the scientists realised exactly how big their discovery was. This giant herbivore lived in the forests of Patagonia between 95 and 100 million years ago, based on the age of the rocks in which its bones were found. There have been many previous contenders for the title "world's biggest dinosaur". | A new species of titanosaur unearthed in Argentina is the largest animal ever to walk the Earth, palaeontologists say. By measuring the length and circumference of the largest femur (thigh bone), they calculated the animal weighed 77 tonnes. "Given the size of these bones, which surpass any of the previously known giant animals, the new dinosaur is the largest animal known that walked on Earth," the researchers told BBC News. "It will be named describing its magnificence and in honour to both the region and the farm owners who alerted us about the discovery," the researchers said. |

Table 1: Sample summaries used in the evaluation

3 Results

We will refer to the two summarisers being compared as NSFC (News Summariser for Children), which uses $Score_{children}$ as the metric and GS (Generic Summariser), which uses $Score_{SumBasic}$. The quantitative data for the four comparison questions are reported in Table 2, with pie charts for each question in Fig. 1.

For statistical analysis of significance, we used the *Sign Test*, by ignoring the 'Not Sure' counts

and aggregating counts for 'slightly' and 'significantly' more. The family significance level was set at $\alpha = 0.05$; with $m = 6$ null hypotheses (that the two summaries are equal on Q1–4 and that for Q5 neither summariser would be considered suitable for children). We used the Bonferroni Correction (α/m), giving an individual significance threshold of $0.05/6 = 0.00833$.

Informative: News Summariser for Children outperformed the generic summariser by a sig-

| | Q1 Info | Q2 Pos | Q3 Easy | Q4 Overall |
|------------------|------------|-----------|------------|---------------|
| GS Significant | 7 | 1 | 7 | 6 |
| GS Slight | 12 | 12 | 9 | 14 |
| Not Sure | 6 | 32 | 14 | 4 |
| NSFC Slight | 26 | 20 | 34 | 32 |
| NSFC Significant | 30 | 16 | 17 | 25 |

Table 2: Responses to comparison questions

nificant margin of 56 to 19 ($p < 0.0001$), with only 14 instances of “Not Sure”. This suggests that the potentially negative effect on informativeness of incorporating sentiment and reading ease into the sentence score was more than offset by our adaptation of the SumBasic score to incorporate increased weight for the first two sentences and ignore stop words.

Positive: While the News Summariser for Children still outperformed the generic summariser by a significant margin of 36 to 13 ($p = 0.0014$), the most common response was “Not Sure”(40% of responses).

Easy: News Summariser for Children outperformed the generic summariser by a significant margin of 51 to 16 ($p < 0.0001$), with only 14 instances of “Not Sure”.

Overall: News Summariser for Children outperformed the generic summariser by a significant margin of 57 to 20 ($p < 0.0001$), with only 4 instances of “Not Sure”.

Non-comparison question: The final question Q5 simply asked the participant to rate whether they would show each summary to a child on a Likert scale. This question was necessary as the News Summariser for Children could have radically outperformed the generic summariser whilst still not have produced a particularly good summary in and of itself. Table 3 presents the quantitative data for the non-comparison question. While the generic summariser (GS) produced output deemed suitable for being shown to children for slightly fewer than half the cases (38 out of 58 where an opinion was expressed; not significant with $p = 0.0124$), the news summariser for children

(NSFC) produced output deemed suitable for the vast majority of cases (69 out of 74 where an opinion was expressed; $p < 0.0001$).

Overall, these results were deemed to be tremendously positive and indicating that the News Summariser for Children has the potential to be an excellent tool in creating news summaries for children. To gain further insights, we also asked an expert in education to provide some qualitative feedback, as reported below.

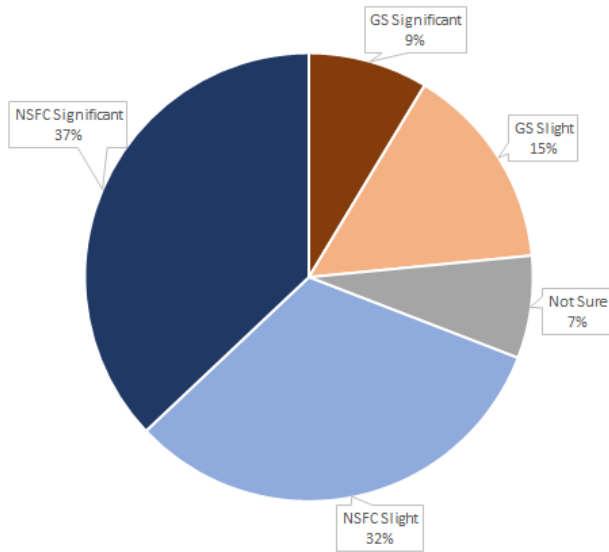
3.1 Qualitative Comparison to BBC Newsround

In order to get qualitative feedback on the strengths and weaknesses of our summariser (NSFC), we selected the summaries of BBC news reports from the previous experiment for which NSFC received the highest and the lowest overall ratings. These were shown to a faculty member from our University’s School of Education, alongside the text from the corresponding BBC Newsround article. The Newsround article and NSFC summary were labelled A or B in each case and no indication was given as to the identity of either. For the NSFC summary rated highest, the qualitative feedback from the expert indicated that the summary created by the NSFC (“B” in the following quote) was actually preferable to the article featured on Newsround (“A” in the following quote):

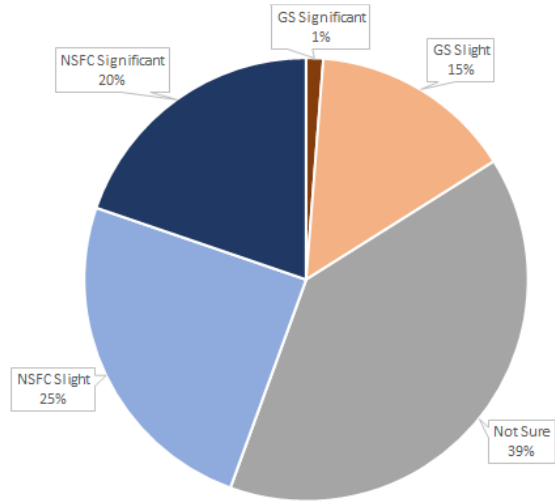
“While A provides more information about the [e]vent, it does not necessarily make the news clearer or understandable. Children up to the age of 10 generally find it difficult to deal with big numbers and with metaphorical words... The level of information [in B] is kept to a minimum. It is very factual and therefore easy to understand

| | GS | NSFC |
|-------------------|----|------|
| Disagree Strongly | 4 | 1 |
| Disagree | 16 | 4 |
| Not Sure | 23 | 7 |
| Agree | 29 | 44 |
| Agree Strongly | 9 | 25 |

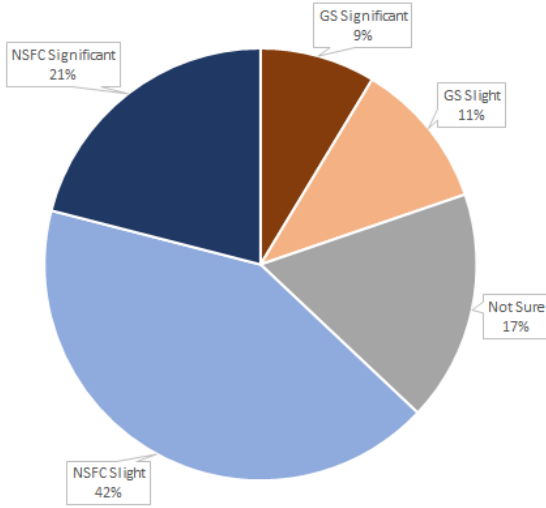
Table 3: Responses to non-comparison question



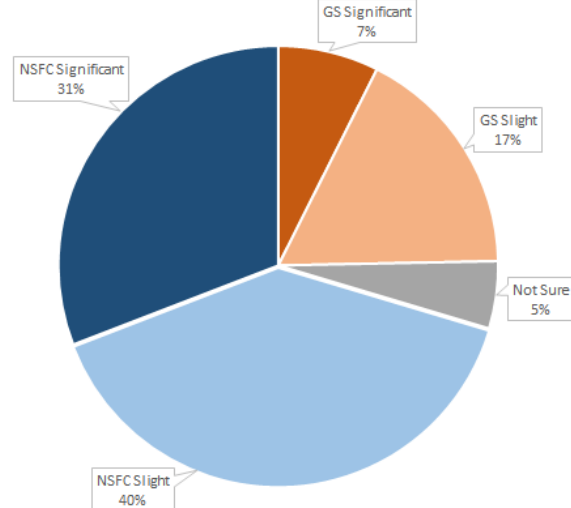
(a) Q1: How informative?



(b) Q2: How positive?



(c) Q3: How easy?



(d) Q4: Overall?

Figure 1: Responses to comparison questions

and to recall. Generally, the usage of common words, short sentences and factual description helps understanding of the news items.”

For the NSFC summary rated lowest, the qualitative feedback from the expert indicated that the summary created by the NSFC (“A” in the following quote) was inferior to the article featured on Newsround (“B” in the following quote):

“A is shorter but ‘denser’ due to the use of scientific jargon, anthropo-

morphised usage of non-human subjects and presence of metaphorical terms.... B is longer and it also includes elements of scientific jargon and metaphorical terms. However the sentences are describing facts effectively by means of clear stating of the subjects, their actions captured by verbs in the active form and places/time”

4 Discussion

While there is considerable work in automatic text summarisation (Nenkova and McKeown,

2012), sentiment analysis (Liu and Zhang, 2012) and computational assessment of text readability (Collins-Thompson, 2014), as well as related fields such as text simplification (Siddharthan, 2014), we are unaware of any work directly targeting the task of summarising news stories for children. Perhaps the most closely related work is De Belder and Moens (2010), who describe a system for *simplifying* news stories in a manner that is suitable for children, splitting sentences up into smaller simpler ones and replacing difficult words with easier synonyms. Related ideas have also been explored in Information Retrieval research, with Collins-Thompson et al. (2011) describing how search results can be reranked by readability to make them suitable for different reading skills, and Enikuomelin and Rahman (2015) describing how sentiment analysis could be incorporated into an IR engine for children.

In the real world, news reporting for children is done manually at considerable cost. The BBC’s CBBC Newsround is a news source with a long history, with its first episode airing in 1972 and regular episodes continuing to broadcast to this day. The primary demographic for these summaries is children aged six to twelve years old (Newsround, 2011). Today a website provides manually written news stories for children. In reality, these stories are often edited versions of an article on the main BBC webpage, but considerably shorter, with easier to read sentences and by and large an optimistic outlook. This is the sort of news story we were attempting to emulate in this paper.

Our quantitative results suggest that our summariser is successful in identifying sentences that are informative while still being upbeat and easy to read. However, there are clearly limitations of our current work. These come through clearly in the qualitative feedback we received from the expert, who made references to “big numbers”, “metaphorical words”, “clear stating of the subjects”, “verbs in the active form”, etc. None of these are captured by our score. Even if they were, it is doubtful whether alternative sentences that are equally informative can be found in a single document summarisation context. The expert also made various spe-

cific observations about vocabulary, highlighting words and phrases such as ‘blaze’, ‘flash floods’, ‘arson’ and ‘aid agencies’ as examples that may be difficult for a child to understand, and approving of Newsround defining terms like ‘arson’ clearly within the text. The solution it would appear is to combine the purely extractive approach described in this paper with more abstractive approaches used in research on text simplification. This will be explored in future work. For instance, numerical simplification (Power and Williams, 2012; Bautista et al., 2011), accurate conversion of passive to active voice (Siddharthan, 2010), sentence shortening to preferentially remove difficult words (Angrosh et al., 2014), lexical simplification (De Belder and Moens, 2010; Yatskar et al., 2010), explanatory descriptions of named entities (Siddharthan et al., 2011), simplifying causality and discourse connectives (Siddharthan, 2003; Siddharthan and Katsos, 2010) and defining terminology (Elhadad, 2006) have all been demonstrated for text simplification systems.

5 Conclusions

Our goal was to create an automatic news summarisation system capable of producing summaries suitable for children by combining scores for sentence informativeness, sentiment and difficulty. Our evaluation confirmed that our summariser outperforms a generic summariser focused only on informativeness in each of the aspects of informativeness, positivity and simplicity. Additionally, an overwhelming majority of experimental participants rated the summaries created by this system as being suitable for being shown to children. An expert in the field of education further confirmed that when the system worked well, the summaries were of a high standard and indeed superior to that created by a professional journalist. The expert also analysed reasons for poor performance of the system on other stories. As discussed in the previous section, there is potential for overcoming these by combining the extractive methods described here with abstractive methods from research on automatic text simplification.

References

- Richard P Adler and Judy Goggin. 2005. What do we mean by “civic engagement”? *Journal of Transformative Education*, 3(3):236–253.
- Mandya Angrosh, Tadashi Nomoto, and Advait Siddharthan. 2014. Lexico-syntactic text simplification and compression with typed dependencies. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, pages 1996–2006, Dublin, Ireland, August. Dublin City University and Association for Computational Linguistics.
- Susana Bautista, Raquel Hervás, Pablo Gervás, Richard Power, and Sandra Williams. 2011. How to make numerical information accessible: Experimental identification of simplification strategies. *Human-Computer Interaction-INTERACT 2011*, pages 57–64.
- Michelle J Boyd and Julie Dobrow. 2010. Media literacy and positive youth development. *Advances in child development and behavior*, 41:251–271.
- Michael Buhrmester, Tracy Kwang, and Samuel D Gosling. 2011. Amazon’s mechanical turk a new source of inexpensive, yet high-quality, data? *Perspectives on psychological science*, 6(1):3–5.
- Kevyn Collins-Thompson, Paul N Bennett, Ryan W White, Sebastian de la Chica, and David Sontag. 2011. Personalizing web search results by reading level. In *Proceedings of the 20th ACM international conference on Information and knowledge management*, pages 403–412. ACM.
- Kevyn Collins-Thompson. 2014. Computational assessment of text readability: A survey of current and future research. *ITL-International Journal of Applied Linguistics*, 165(2):97–135.
- Edgar Dale and Jeanne S Chall. 1948. A formula for predicting readability: Instructions. *Educational research bulletin*, pages 37–54.
- Jan De Belder and Marie-Francine Moens. 2010. Text simplification for children. In *Proceedings of the SIGIR workshop on accessible search systems*, pages 19–26. ACM.
- Noemie Elhadad. 2006. Comprehending technical texts: Predicting and defining unfamiliar terms. In *AMIA Annual Symposium proceedings*, volume 2006, page 239. American Medical Informatics Association.
- AO Enikuomelin and MA Rahman. 2015. A quantum based web summarizer for children’s news rendering. *International Journal of Internet of Things*, 4(1):6–10.
- Ken J Gilhooly and Robert H Logie. 1980. Age-of-acquisition, imagery, concreteness, familiarity, and ambiguity measures for 1,944 words. *Behavior Research Methods & Instrumentation*, 12(4):395–427.
- Derek Greene and Pádraig Cunningham. 2006. Practical solutions to the problem of diagonal dominance in kernel document clustering. In *Proc. 23rd International Conference on Machine Learning (ICML’06)*, pages 377–384. ACM Press.
- Robert Gunning. 1952. The technique of clear writing.
- Matthias Hüning. 2002. Textstat. <http://neon.niederlandistik.fu-berlin.de/static/textstat/TextSTAT-Doku-EN.html#7>.
- JP Kincaid, RP Fishburne, RL Rogers, and BS Chissom. 1975. Derivation of new readability formulas. Technical report, Technical report, TN: Naval Technical Training, US Naval Air Station, Memphis, TN.
- Bing Liu and Lei Zhang. 2012. A survey of opinion mining and sentiment analysis. In *Mining text data*, pages 415–463. Springer.
- Bing Liu, Mingqing Hu, and Junsheng Cheng. 2005. Opinion observer: analyzing and comparing opinions on the web. In *Proceedings of the 14th international conference on World Wide Web*, pages 342–351. ACM.
- Hans Peter Luhn. 1958. A business intelligence system. *IBM Journal of Research and Development*, 2(4):314–319.
- G Harry Mc Laughlin. 1969. Smog grading—a new readability formula. *Journal of reading*, 12(8):639–646.
- Ani Nenkova and Kathleen McKeown. 2012. A survey of text summarization techniques. In *Mining Text Data*, pages 43–76. Springer.
- Ani Nenkova and Lucy Vanderwende. 2005. The impact of frequency on summarization. *Microsoft Research, Redmond, Washington, Tech. Rep. MSR-TR-2005-101*.
- CBBC Newsround. 2008. Cbbc newsround - “who we are”. <http://news.bbc.co.uk/cbbcnews/hi/help/default.stm>. Accessed: 16th November, 2015.
- CBBC Newsround. 2011. Cbbc newsround - frequently asked questions. <http://www.bbc.co.uk/newsround/13927399>. Accessed: 13th April, 2016.
- Bo Pang and Lillian Lee. 2004. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In *Proceedings of the ACL*.
- Horst Pöttker. 2003. News and its communicative quality: the inverted pyramid—when and why did it appear? *Journalism Studies*, 4(4):501–511.

- Richard Power and Sandra Williams. 2012. Generating numerical approximations. *Computational Linguistics*, 38(1):113–134.
- Sarah R Ramsey, Kristen L Thompson, Melissa McKenzie, and Alan Rosenbaum. 2016. Psychological research in the internet age: The quality of web-based data. *Computers in Human Behavior*, 58:354–360.
- David G Rand. 2012. The promise of mechanical turk: How online labor markets can help theorists run behavioral experiments. *Journal of theoretical biology*, 299:172–179.
- Carole Rich. 2015. *Writing and reporting news: A coaching method*. Cengage Learning, eighth edition.
- Advaith Siddharthan and Napoleon Katsos. 2010. Reformulating discourse connectives for non-expert readers. In *Proceedings of the 11th Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT 2010)*, Los Angeles, CA.
- Advaith Siddharthan, Ani Nenkova, and Kathleen McKeown. 2011. Information status distinctions and referring expressions: An empirical study of references to people in news summaries. *Computational Linguistics*, 37(4):811–842.
- Advaith Siddharthan. 2003. Preserving discourse structure when simplifying text. In *Proceedings of the European Natural Language Generation Workshop (ENLG), 11th Conference of the European Chapter of the Association for Computational Linguistics (EACL’03)*, pages 103–110, Budapest, Hungary.
- Advaith Siddharthan. 2010. Complex lexico-syntactic reformulation of sentences using typed dependency representations. In *Proceedings of the 6th International Natural Language Generation Conference*, pages 125–133. Association for Computational Linguistics.
- Advaith Siddharthan. 2014. A survey of research on text simplification. *ITL-International Journal of Applied Linguistics*, 165(2):259–298.
- Hans Stadthagen-Gonzalez and Colin J Davis. 2006. The bristol norms for age of acquisition, imageability, and familiarity. *Behavior research methods*, 38(4):598–605.
- University of Washington. 2012. Stopword.list. <http://courses.cs.washington.edu/courses/cse573/12sp/A2/>.
- Grady Ward. 1993. Names.txt. <http://www.gutenberg.org/ebooks/3201>.
- Martin Weisser. 2013. Iceweb. http://martinweisser.org/ling_soft.html#iceweb.
- Mark Yatskar, Bo Pang, Cristian Danescu-Niculescu-Mizil, and Lillian Lee. 2010. For the sake of simplicity: Unsupervised extraction of lexical simplifications from wikipedia. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 365–368. Association for Computational Linguistics.